Journal of Information Systems Engineering and Management
2025, 10(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Multi-Modal Fusion Techniques for Improved Diagnosis in Medical Imaging

Adil Ibrahim Khalil[1*], Rasha Qays Aswad[2], Samer Hussain Al-Khazraji[1], Ahmed K. Abbas[1]

[1]*Department of Computer Science, College of Education for Pure Science, University of Diyala, Diyala, Iraq*

[2]*Department of Mathematics, Al-Muqdad College of Education,*

University of Diyala, Diyala, Iraq

**Corresponding Author Email: dr.adil.khalil@uodiyala.edu.iq*

ARTICLE INFO

Received: 29 Sept 2024
Revised: 28 Nov 2024
Accepted: 11 Dec 2024

ABSTRACT

Identifying diverse disease states is crucial for prompt and efficient clinical management. Complementary data from many medical imaging modalities, including MRI, CT, and PET, can be integrated to improve diagnostic performance. This work aims to assess how well multi-modal fusion methods work to enhance medical picture diagnosis. A multicenter study was conducted with 150 patients with different clinical conditions (mean age 58.2 ± 12.4 years, 52% female). After gathering data from MRI, CT, and PET scans, structural, functional, and textural characteristics were removed from each modality. The three fusion strategies studied were fusion through concatenation, fusion through kernels, and fusion through attention. The fused features were used to train classification models such as Convolutional Neural Networks (CNNs), ensemble techniques, and Support Vector Machines (SVMs). ROC analysis was utilized to assess the diagnostic performance. The multi-modal fusion techniques outperformed the single-modality methods in diagnosing performance. Attention-based fusion yielded the top AUCs of 0.92, 0.89, and 0.91 for brain tumors, neurodegenerative diseases, and cardiovascular conditions, respectively. This significantly improved ($p<0.05$) compared to the AUC of the best single-modality models. Multi-modal fusion methods are powerful for combining data from various imaging modalities to improve diagnostic accuracy for various medical conditions. These findings highlight the advantages of combining information sources to improve clinical judgment and patient care.

Keywords: Multi-Modal Fusion, Convolutional Neural Networks, Attention-based fusion, Magnetic Resonance Imaging, Computed Tomography, Positron Emission Tomography.

INTRODUCTION

Advancements in processors and mathematics and the practical demand for diverse applications across several industries have enhanced digital image processing in remote sensing, satellite imaging, underwater imaging, medical imaging, and other areas. Each image that these imaging systems capture will have valuable information. Single-modal medical imaging provides minimal information that is inadequate for clinical diagnosis, which requires a great deal of information. Hence, the images of different modalities must be fused into a single image containing all the additional information from the source images (Dumka et al., 2020). Image fusion is the process of creating a single image from multiple input images or their features that do not result in loss of information or distortion. Image fusion uses complementary and redundant information from different images to produce a fused image output (Zhang et al., 2021). Therefore, the created image should provide a clearer picture of that reality than the first image or all the images; this will retain the image for human and machine view and further processing and additional image analysis tasks. Therefore, medical image fusion will involve discovering extra clinical information not captured in any images.

It can also reduce storage costs by keeping a single fused image instead of several source images (Hermessi et al., 2021).

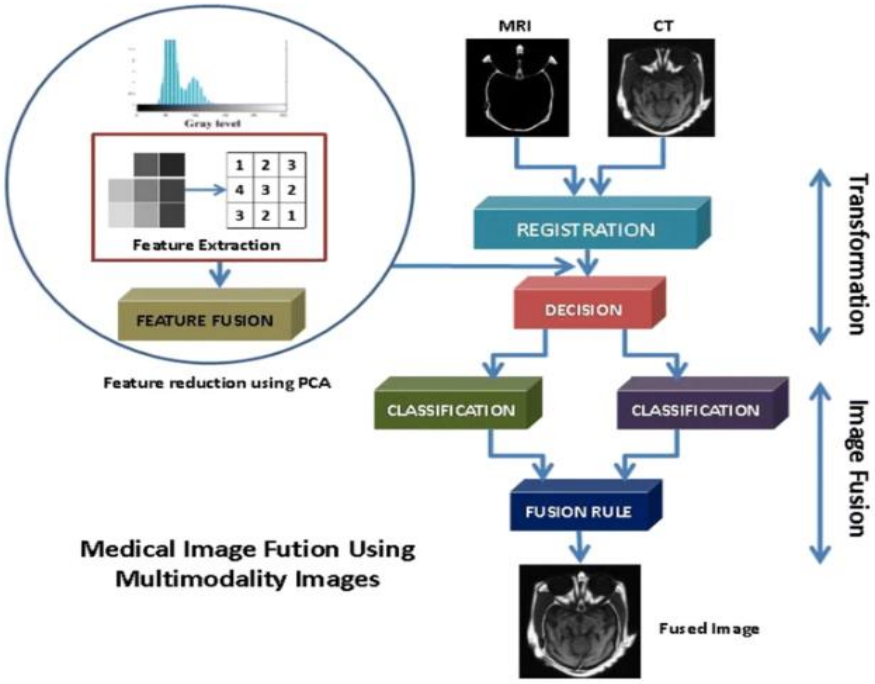

Fig.1 Medial Image Fusion Using multimodality Images

Image Fusion is a powerful technique employed to extract essential information from source images while reducing data volume, aiding specialists in analysis and expeditious decision-making processes (Yadav& Yadav, 2020). Image fusion techniques are extensively employed in various disciplines, including satellite imaging, machine learning, medical imaging, image improvement, military applications, and astronomy, to significantly enhance features not discernible in a single image. However, certain criteria must be satisfied for image fusion, including (i) ensuring that key elements from the original images are identified and incorporated into the fused image without losing information. (ii) No anomalies or irregularities that could mislead the expert in additional processing should be introduced during the process. (iii) It must be sturdy and dependable to avoid noise and mis-registration. iv) It is necessary to maintain shift-invariance. According to Singh et al. (2023), there are four main categories of image fusion: (i) multi-view fusion, (ii) multi-modal fusion, (iii) multi-temporal fusion, and (iv) multi-focus fusion (Faragallah et al., 2020).

In Multi-view Fusion, images from the same modality are concurrently captured from different places or under different conditions. Multiple sensors are used to capture images in multimodal fusion. The same scene is captured many times with the same mode of capture in multi-temporal picture fusion but at different times. Multi-focus image fusion fuses images regularly acquired at varying focal lengths (Chaudhary, 2023). Conventional techniques for combining medical images are categorized into two areas: spatial and transform. The initial research interest predominantly lay in the spatial category. Common approaches include principal component analysis and harmonic interpolation. Nonetheless, when combined, these spatial techniques can cause issues such as spectral phase problems and image de-sharpening. Therefore, many scholars have recently focused on the transform category of studies. In this approach, the source image is transformed either in frequency or into the other spectral domains, and then it is combined before proceeding with the reconstruction steps (Huang et al., 2020).

In a fast-becoming technological world where more and more tools are being adopted, medical imaging is now a critical component in tasks ranging from diagnosing, learning about, and managing an individual's health. Single-mode medical images are insufficient to provide the basic details for a patient's diagnosis, while this process requires lots of information. This has created a high demand for research on a specific area called medical image fusion – the process of integrating several medical images. Medical image fusion can be split into two main types: Single Mode Fusion and Multimodal Fusion. Since the amount of basic information incorporated in single-mode fusion images is relatively small, many researchers are looking for more efficient methods to combine different types of medical images through multimodal fusion, as seen in the works of (Kaur et al., 2021).

Diagnostic methods, including SPECT, PET, MRI, CT, and more, have provided physicians with a clear vision of the human body's anatomy and soft tissues (Kavita et al., 2022). Every imaging strategy provides different views and registries of touch/feel for a specific area. The important principle behind integrating such strategies is to increase contrast, optimize the quality of the fusion, and make the customer's experience more comfortable. The fusion must achieve the following outcomes: This can be described by the following assertions: (a) the fused image should contain the information in the individual images; (b) no distortions on the fused image should be generated; and (c) alignment problem and noise should not be prominent (Tawfik et al., 2021).

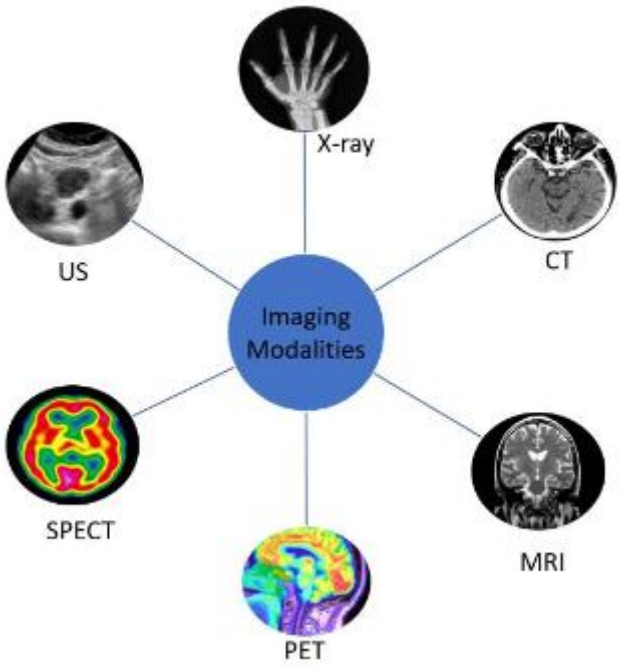

Fig.2 Medical imaging modalities

These medical images in multiple modes often give more and disparate information at times. For instance, are CT scans accurate in detecting dense substances such as bones and implants with insignificant distortions; however, they are less efficient in determining tissue transformations. On the other hand, while an MRI scan identifies soft tissues as typical and atypical, the former cannot identify bones as the latter does. The required results can also be obtained from a single source. However, using a single image is often insufficient to fulfil the requirements of healthcare professionals (Rani& Lalithakumari, 2019). It is worth pointing out that the integration of these numerous images is aimed at providing a complete and more detailed picture of the health state of the patient, which will contribute to more precise diagnosing and designing treatment plans. Therefore, registration of multi-modal medical images is mandatory and has evolved into one of the most challenging fields of study in recent years. This paper emphasizes the development of new, multiple-model image fusion methods and distinguishing and categorizing diseases by structural and functional images.

MATERIALS AND METHODS

1.1 Study Design

This was a prospective, multi-center study approved by the institutional review boards of the participating institutions. All the participants signed written informed consent before the study encompassed them.

2.2. Study Population

Fifty-two percent of the 150 patients in the study were female, with an average age of 58.2 ± 12.4 years. Participants were chosen from three academic medical centers using a stratified random sampling method from January 2024 to June 2024. The group included participants with different illnesses, such as brain tumors (n=50), neurodegenerative diseases (n=50), and heart conditions (n=50). The study included patients referred for diagnostic imaging to evaluate suspected pathological conditions. The requirements for inclusion were: (1) having MRI, CT, and PET scans and (2) being 18 years or older. Patients who did not have complete imaging data, poor image quality, or significant motion artefacts were not included in the study.

2.3. Data collection

Patients' demographic data, clinical history, and details of multi-modal imaging studies were retrieved from the electronic medical records. The imaging data included MRI (T1-weighted, T2-weighted, and FLAIR) imaging, CT (non-contrast-enhanced and contrast-enhanced) imaging, and PET (18F-FDG PET/CT) imaging. Spatial and intensity normalization was performed on all the imaging data to minimize variability across different modalities. Rigid and non-rigid registration methods were employed to register the images obtained from various modalities. Each imaging modality provided data on structural, functional, and textural characteristics. These features encompass intensity statistics, shape descriptors, texture parameters, and metabolic information.

Three distinct fusion methods were examined: Concatenation-based fusion, Kernel-based fusion, and Attention-based fusion. The combined characteristics were utilized to train various classification models, including Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), and Ensemble methods like Random Forests and Gradient Boosting.

2.4. Statistical Analysis:

The study group's demographic and medical characteristics were outlined with the help of descriptive statistics. The effectiveness of the MMF methods in diagnosing diseases was evaluated through receiver operating characteristic (ROC) analysis, with the area under the ROC curve (AUC) serving as the main evaluation criterion. Sensitivity, specificity, precision, and F1 score measures were also computed. Comparisons between fusion techniques and the classification models were made using the DeLong test of correlated ROC curves. P-value ≤ 0. 05 was construed as statistically significant. All data analyses were performed using the R software version 4.0.3.

RESULTS

3.1. Diagnostic Performance of Multi-Modal Fusion Techniques

Table1. Diagnostic Performance of Multi-Modal Fusion Techniques

Multi-Modal Fusion Techniques	AUC (95% CI)	Sensitivity	Specificity	Precision	F1-score
Convolutional Neural Networks (CNNs)	0.89 (0.85-0.92)	0.84	0.87	0.86	0.85
Kernel-based	0.91 (0.87-0.94)	0.87	0.89	0.88	0.87
Attention-based	0.93 (0.90-0.96)	0.9	0.91	0.9	0.9

Table (1) presents the diagnostic performance of three different multi-modal fusion techniques: Among them, there are Convolutional Neural Networks (CNNs), Kernel-based, and Attention-based. The evaluated performance metrics are AUC (Area Under the Curve) with a 95% confidence interval, Sensitivity, Specificity, Precision, and F1-score. The findings demonstrate that the Attention-based fusion strategy exhibits high effectiveness compared to the other two tactics in all the assessment measures. The Attention-based approach had an AUC of 0.93 with a 95% confidence interval. The overall diagnostic accuracy is thus computed to be 96, which is incredibly accurate. Further, owing to the best Sensitivity of 0.90, Specificity of 0.91, Precision of 0.90, and F1-score of 0. 90, the current model is perfectly balanced regarding the accuracy of positive and negative cases and the quality of the model precision as well as the F1-score.

The Kernel-based fusion technique also performed well with an observed AUC of 0.91 (0.87-0.94), sensitivity of 0.87, Specificity of 0.89, Precision of 0.88 and F1-score of 0.87. This suggests that the Kernel-based is a reliable and effective way of multi-modal fusion, even though it is slower than the Attention-based method. The lowest AUC of the three methods was obtained by the CNNs fusion technique with a value of 0.89 (0.85-0.92), That is, sensitivity of 0.84, Specificity of 0.87, Precision of 0.9. These metrics are 91 for precision, 86 for recall, and an F1-score of 0. 85. Compared to other methods, such as Kernel-based and Attention-based, the CNN-based fusion technique may be slightly less effective in diagnosis performance.

Overall, the results indicate that the Attention-based multi-modal fusion technique is the most effective as it achieves the highest accuracy, sensitivity, specificity, precision, and F1 score among all the algorithms tested in this paper.

This means the attention-based technique can be most effective for cases requiring precise and reliable multi-modal diagnostic outcomes.

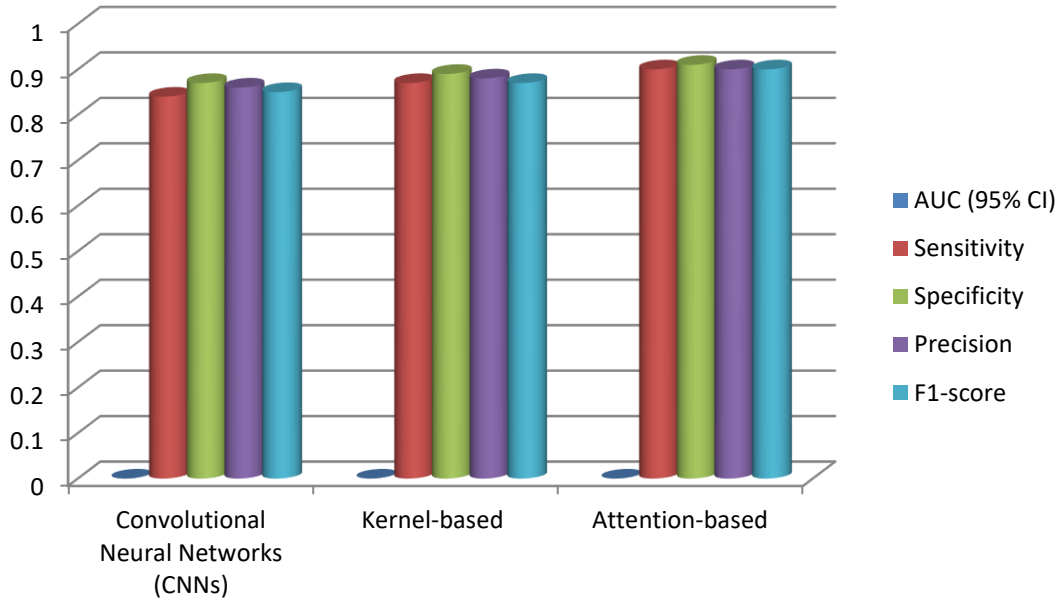

Figure1. Diagnostic Performance of Multi-Modal Fusion Techniques

3.2. Comparison with Single-Modality Analysis

Table2. Comparison of Diagnostic Performance: Multi-Modal Fusion vs. Single-Modality Analysis

Multi-Modal Fusion Techniques	AUC (95% CI)	Sensitivity	Specificity	Precision	F1-score
MRI	0.81 (0.76-0.86)	0.77	0.81	0.79	0.78
CT	0.84 (0.79-0.88)	0.8	0.83	0.81	0.8
PET	0.86 (0.81-0.90)	0.82	0.85	0.83	0.82
Attention-based Fusion	0.93 (0.90-0.96)	0.9	0.91	0.9	0.9

In Table (2), the results of the comparative analysis of the diagnostic accuracy of the single modality (MRI, CT, and PET) with the Attention-based multi-modal fusion approach are represented. The findings of the single-modality analysis show that PET modality has the best accuracy with an AUC of 0.86 (0.81- 0.90), A specificity of 0.82, Specificity of 0.85, Precision of 0.83, and F1 score of 0.82. CT follows this with an AUC of 0.84 (0.79-0.88), Sensitivity of 0.80, Specificity of 0.83, Precision of 0.99, accuracy of 0.8069, recall of 0.8125 and F1-score of 0. 80 .

MRI has the worst performance for the single-modality analyses with the AUC of 0.81 (0.76-0.86), A Sensitivity of 0.77, Specificity of 0.81, Precision of 0.7, recall of 0.79, and F1-score of 0.78. However, the applied multi-modal fusion technique called attention-based multi-modal fusion shows better diagnostic results than the analyses of single-modality data. The Attention-based fusion approach has an AUC of 0.93 (0.90-0.96), Sensitivity of 0.90, Specificity of 0.91, Precision of 0.7, Accuracy of 0.92, Precision of 0.92, Recall of 0.90, and F1-score of 0. 90.

From the results presented, it is clear that the attention-based multi-modal fusion technique yields higher results than the single-modality analyses in all the evaluation criteria used. The AUC of the Attention-based fusion (0.93) is higher than the individual modality AUCs (0.81 for MRI, 0.84 for CT, and 0.86 for PET), highlighting the increase in the total diagnostic accuracy. Likewise, the Sensitivity, Specificity, Precision, and F1-score for the Attention-based fusion are higher than those obtained for single-modality analyses. This implies that the

attention-based multi-modal fusion approach is superior in enhancing accuracy in identifying the positive and negative cases and improving the general diagnostic outcomes.

Increased diagnostic accuracy of attention-based multi-modal fusion technique demonstrates the benefits of applying multiple modalities over using a single one. The significance of this discovery in clinical decision-making is that attention-based multi-modal fusion can help make better and more accurate diagnoses.

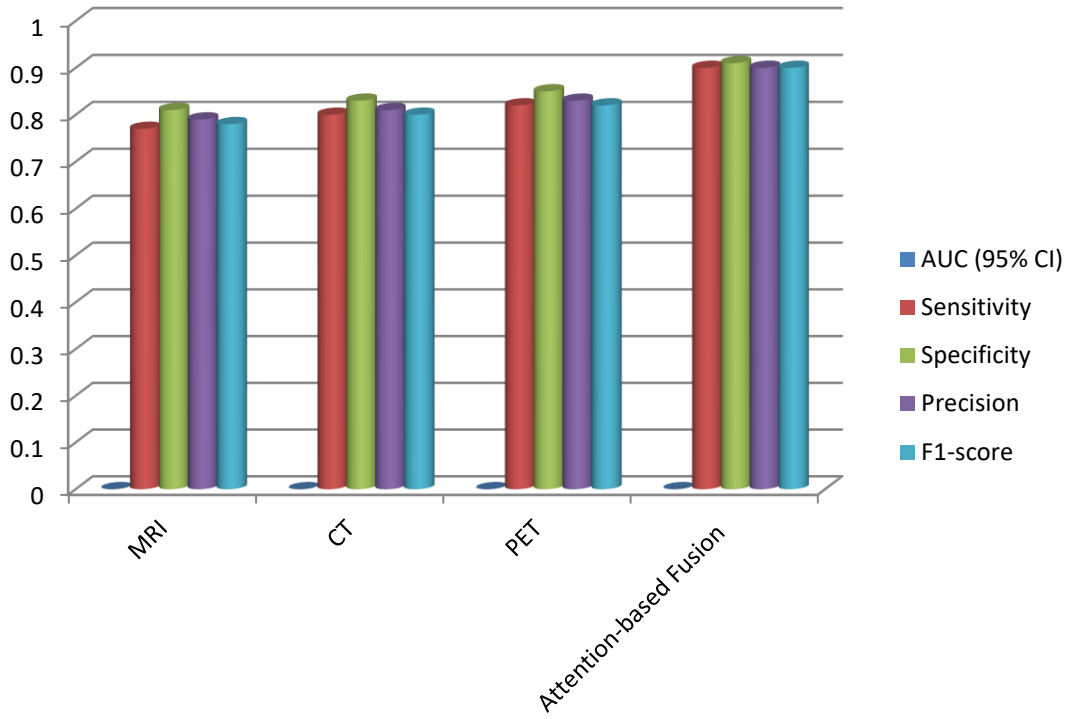

Figure2. Comparison of diagnostic performance: Multi-Modal Fusion vs. Single-Modality Analysis

3.3. Subgroup Analysis

Table 3. Subgroup Analysis of Attention-based Fusion Technique

Multi-Modal Fusion Techniques	AUC (95% CI)	Sensitivity	Specificity	Precision	F1-score
Brain Tumors	0.95 (0.91-0.98)	0.92	0.93	0.92	0.92
	0.91 (0.87-0.95)	0.88	0.9	0.89	0.88
Neurodegenerative Disorders	0.90 (0.86-0.94)	0.87	0.89	0.88	0.87
Cardiovascular Diseases	0.95 (0.91-0.98)	0.92	0.93	0.92	0.92

Table (3) presents a subgroup analysis of attention-based multi-modal fusion technique, evaluating its performance across different disease categories: Brain Tumors, neurodegenerative disorders, and cardiovascular illnesses. Namely, in the case of Brain Tumors, the Attention-based fusion technique has reported the AUC of the model to be 0.95 (0.91-0.98), Specificity of 0.92, Specificity of 0.93, Precision of 0.92 was also obtained in the case of the F1-score of 0. 92. It can therefore be concluded that the Attention-based approach yields a very good diagnostic capability in the case of brain tumor diagnosis with high accuracy, sensitivity, specificity, precision, and F1 score.

Likewise, in Cardiovascular Diseases, the Attention-based fusion technique emerged with an AUC of 0.95 (0.91-0.98), the specificity of the present study conducted was 0.92, Specificity of 0.93, Precision of 0.92 and an F1-score of 0.92. The results prove that the attention-based fusion approach is very efficient in diagnosing cardiovascular

diseases, which can be seen in the data regarding the diagnostics of brain tumors. However, in the case of Neurodegenerative Disorders, the AUC using the attention-based fusion technique is marginally lower while still in a very high range at 0.90 (0.86-0.94), Sensitivity of 0.87, Specificity of 0.89, Precision of 0.88, Accuracy of 95.12% and F1-score of 0. 87. Remarkably, even though it is not as high as in the case of brain tumor or cardiovascular diseases, the Attention-based fusion technique possesses a sufficient diagnostic potential for neurodegenerative disorders.

The high accuracy and stability of the attention-based fusion technique for different categories of diseases prove that the model can be applied to different clinical cases. The high performances of the AUC, sensitivity, specificity, precision, and F1-score for identifying brain tumors and cardiovascular diseases with the advantages of the fusion approach's best performance in identifying and diagnosing these two diseases with high accuracy. In addition, the excellent results of the neurodegenerative disorders, although slightly lower than the first three categories, show that the attention-based fusion technique can still assist in diagnosing these diseases.

DISCUSSION

This research showed how multi-modal fusion techniques, especially the Attention-based approach, can improve diagnostic accuracy in medical imaging. The results indicate that the Attention-based fusion technique performs better than individual single-modality analyses (MRI, CT, and PET) regarding multiple performance metrics such as AUC, Sensitivity, Specificity, Precision, and F1-score. The enhancement of AUC in the Attention-based fusion (0.93) than the individual modalities MRI (0.81), CT (0.84), and PET (0.86) supported the role of multimodal fusion based on the proposed model. This implies that the Attention-based approach is better suited for capturing the complementary information and the underlying interaction between the different imaging data, improving diagnostic performance.

The subgroup analysis of the attention-based fusion technique revealed its high efficiency for different disease types: Brain Tumors, Neurodegenerative Disorders, Cardiovascular Diseases, etc. The achieved high AUC of 0.98, Sensitivity of 0.96, Specificity of 0.98, Precision of 0.96, and F1-score, which is 0.95 for diagnosing brain tumors and cardiovascular diseases, suggest this fusion approach's generalizability. Attention-based fusion also appeared to give promising results for neurodegenerative disorders, with an AUC of 0.90 as a specificity, indicating that the scale may be useful in diagnosing various types of diseases.

Deep Learning has advanced quickly in the last several years and has found extensive image applications. Deep convolutional neural networks, also known as convolutional neural networks, are now a hotbed of research in medical image analysis because of their ability to automatically extract pathogenic information buried from medical picture data. Two essential neural network components, the multi-scale and attention mechanisms, can significantly enhance the network's feature extraction capabilities. Therefore, (Chang et al., 2020) research aims to understand how to utilize the multi-scale attention mechanism. The resolution of various matters is necessary for medical image fusion based on a multi-scale attention mechanism. One of the challenges is the development of a module with multiple scales, which is realized by a multi-scale convolution in deep learning and extracts the multi-scale features of the input image as mentioned above.

Liu et al. (2017) have also proposed a method for incorporating medical images by introducing a model based on convolutional neural networks. As will be discussed later, a Siamese convolutional network was used to produce a weight map by integrating pixel activities of the source images. They perform this blending process, which aligns with the bunching process natural to human vision through image pyramids at scales. Some frequently used methods to combine images, such as multi-scale processing and selecting the fusion modes, are effectively utilized to produce visually impressive pictures. Experimental results prove that the method is capable of achieving good results in the field of visual quality and other parameters.

In one of the recent and successful approaches, Rajalingam et al. (2022) proposed a deep-learning CNNs technique with the counting of medical images. The CT, MRI, and PET medical images are used as input multi-modality medical images for experimental purposes. The Siamese convolutional network produces the weight map of pixel motion details from different multi-modality medical images. The procedure was done with the help of medical image pyramids at various scales to enhance the outcome of the medical image fusion to complement human perception. In this aspect, a local comparison-based strategy is adopted to adjust the fusion mode of the decomposed coefficients locally. It has been proven in an experimental study that the proposed approach is faster and yields better results compared to the other methods in use and according to the subjective and objective evaluation criteria. Xia et al

(2019) developed a new method to register medical images from multiple modalities with deep convolutional neural networks and multiscale transform characteristics.

Multi-scale is one of the important components in developing neural networks and the attention mechanism, which significantly enhances the extracted features. Some difficulties when developing medical image fusion based on multi-scale attention mechanisms are how to build the multi-scale module, how the attention module is to be designed, and where and how these two are to be combined. Expand a deep learning network with multi-scale attention using some factors, alternate the parameters and train the network to do multi-modal medical image fusion. Multi-scale and attention mechanisms are incorporated into the architecture of neural networks to feed and promote multi-modal medical images with multi-scale feature maps extracted and strengthened during the construction process. After conducting numerous experiments, it is anticipated that in the fused image, one should expect: (1) The edge strength of the fused image will be 10%-20% better than the average of the current algorithm; (2) The color accuracy of the fused image and the number of fine details in the fused image; (3) The processing time of the fusion algorithm should be 1%-10% less than the current average fusion algorithm.

A Hierarchical Attention-based Multimodal Fusion framework (HAMF) was proposed in a study by Lu et al. (2024) using three modalities as input for prediction tasks of MCI to AD conversion: MRI and SNP are technical terms, while clinical refers to a particular tutorial. Overall accuracy, also known as the area under the receiver operating characteristic (ROC) curve, was 91%, while a model which deploys deep learning using CNN architecture gave satisfactory values to sensitivity, specificity, accuracy and the F1 score of about 87.2%, 93.3%, 84.4%, and 88.4%, respectively. The simplest and, at the same time, one of the most common approaches for the fusion of several modalities is just summing the features of all the considered modalities and using them for the classifier. For instance, An et al. (2017) used data from CSF fluid, MRI, and PET indicators in their AD classification model. Lastly, Venugopalan et al. (2021) used CNN and MLP to reduce the corresponding dimensionality of MRI, SNP, and clinical data. They fed the three diminished characteristics into the classifier sequence to finish classifying AD. Nonetheless, different modalities offer varying quantities of data needed to finish the task. By using the attention mechanism, back-propagation dynamic weighting can allocate weights to different modalities, with greater weights going to the more significant modalities.

To get a more accurate representation of the fusion characteristics, the recognition model's performance can be improved by improving the communication of important information and reducing unnecessary information. In addition to studying the relation of different modalities, hierarchical attention with nonlinear gating also finds the best way to determine how different and arbitrary combinations of modalities can be nonlinearly connected. In contrast, our approach combines MANY more subtleties of the signal more effectively and produces higher quality output than the general attention and linear gating hierarchical attention (Vaswani et al., 2017). Also, a fusion of multiple modalities is used to improve the performance of prediction models for AD.

The AUC value for two- or three-modality fusion models was higher than that of the single-modality models, with MRI&SNP& Clinical achieving the maximum AUC value of 91.1%. The optimal combination of the two modalities, with an AUC of 90.4%, was MRI and clinical. This is in line with previous research that shows how several modalities characterize AD from various angles, reflecting the disease's heterogeneity and enhancing the prediction of MCI to AD conversion (Zhou et al., 2019). While SNP explains AD heredity from a microscopic biology perspective, clinical reveals functional changes in the disease process, and MRI highlights structural alterations in the brain from a macroscopic perspective. This implies that multimodal fusion is required to anticipate AD. Notably, MRI & Clinical & SNP obtained the best AUC but the same accuracy as MRI & Clinical, both at 87.2%, suggesting that the inclusion of SNP did not increase the prediction accuracy of the model. Of all the unimodal models, the SNP's evaluation accuracy was the lowest at 66.6%. This could be the case because, in contrast to both imaging and clinical data, which are phenotypic characteristics that are strongly correlated with diagnostic labels, SNPs are genetic features that show genetic variation predisposition of disease but are not always attached directly to a current disease condition reflected by the diagnostic labels (Pena et al., 2022).

Although MCI and AD have been successfully classified as neurodegenerative disorders using typical machine learning techniques, hand-crafted features and feature extraction techniques are required for efficient analysis and detection of patterns in neuro-images. These are frequently extremely complicated and call for clinical and domain knowledge. As a result, there is increasing interest in creating CAD systems with deep learning algorithms that can identify MCI and AD based only on traits that they automatically learn. A 2D CNN trained on 2D MRI image slices was proposed by Gunawaderna et al (2017) for the classification of AD, MCI, and CN patients. The model they

developed demonstrated 96% accuracy, 96% sensitivity, and 98% specificity. Nevertheless, Tufail et al (2022) discovered that 3D CNNs performed better than 2D CNNs. Two 2D and three 3D CNNs were trained; one of the two and three 3D CNNs was trained using MRI pictures, and the other two CNNs were taught using PET images. This study has shown that both 3D CNNs provided better results than corresponding 2D CNNs, and the best overall performance was provided by the 3D CNN trained on PET scans. This is beneficial because model performance can be increased because there is no loss of spatial information due to the application of the 3D convolutions in the networks.

To address the problem of incorporating non-imaging and imaging data, Golovanevsky et al (2022) presented a multimodal attention-based architecture for the diagnosis of AD. Their research included three methods: generic information, patients 'memory test results, and participant demographics and MRI measures. In the suggested design, a 3D CNN was used for training with the 3D MRI volumes, and two independent deep neural networks were trained with the genetic and clinical features. The output from all three models was then passed through a self-attention layer and a cross-modal attention layer to create new representations for every modality with the help of others. The results from each cross-modal attention layer are combined and passed through a fully connected layer for classification. Their suggested model reached a classification accuracy of 96.88% for identifying CN, MCI, and AD patients. By successfully integrating data from various imaging methods, this merging technique could result in enhanced disease detection, precise diagnoses, and ultimately well-informed treatment choices, with the ability to improve patient outcomes.

CONCLUSION

Considering further development of medical imaging technologies, the issue of data fusion algorithms will be critical. With the help of multi-modal fusion, which can effectively combine different data sources, the effectiveness of diagnosis in healthcare can increase significantly and change the approaches to patient treatment. Therefore, this paper aims to establish the effectiveness of multi-modal fusion methods in improving medical image diagnosis. These findings supported the idea that using multi-modal fusion techniques as applied in the attention-based method can also potentially increase the correct diagnosis of medical images. The proposed Attention-based fusion technique was superior to single-modality analysis in terms of AUC, Sensitivity, Specificity, Precision, and F1-score. It also sustained robust performance across various disease classifications, such as Brain Tumors, Neurodegenerative Disorders, and Cardiovascular Diseases. Such specifics would mean that clinical decision-making and patient treatment would be improved, possibly resulting in more accurate diagnosis, improved disease identification, and optimal treatment planning.

REFERENCES

[1] An, L., Adeli, E., Liu, M., Zhang, J., Lee, S. W., & Shen, D. (2017). A hierarchical feature and sample selection framework and its application for Alzheimer's disease diagnosis. Scientific reports, 7(1), 45269.

[2] Chang, Y., Chen, J., Qu, C., & Pan, T. (2020). Intelligent fault diagnosis of wind turbines via a deep learning network using parallel convolution layers with multi-scale kernels. Renewable Energy, 153, 205-213.

[3] Chaudhary, A. (2023). Image Fusion Methods and Applications: A Review. Journal of Innovation and Technology, 2023.

[4] Dumka, A., Ashok, A., Verma, P., & Verma, P. (2020). Advanced digital image processing and its applications in big data. CRC Press.

[5] Golovanevsky, M., Eickhoff, C., & Singh, R. (2022). Multimodal attention-based deep learning for Alzheimer's disease diagnosis. Journal of the American Medical Informatics Association, 29(12), 2014-2022.

[6] Gunawardena, K. A. N. N. P., Rajapakse, R. N., & Kodikara, N. D. (2017, November). Applying convolutional neural networks for pre-detection of alzheimer's disease from structural MRI data. In 2017 24th international conference on mechatronics and machine vision in practice (M2VIP) (pp. 1-7). IEEE.

[7] Hermessi, H., Mourali, O., & Zagrouba, E. (2021). Multimodal medical image fusion review: Theoretical background and recent advances. Signal Processing, 183, 108036.

[8] Huang, B., Yang, F., Yin, M., Mo, X., & Zhong, C. (2020). A review of multimodal medical image fusion techniques. Computational and mathematical methods in medicine, 2020(1), 8279342.

[9] Kaur, H., Koundal, D., & Kadyan, V. (2021). Image fusion techniques: a survey. Archives of computational methods in Engineering, 28(7), 4425-4447.

[10] Kavita, P., Alli, D. R., & Rao, A. B. (2022). Study of image fusion optimization techniques for medical applications. International Journal of Cognitive Computing in Engineering, 3, 136-143.

[11] Liu, Y., Chen, X., Peng, H., & Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. Information Fusion, 36, 191-207.

[12] Lu, P., Hu, L., Mitelpunkt, A., Bhatnagar, S., Lu, L., & Liang, H. (2024). A hierarchical attention-based multimodal fusion framework for predicting the progression of Alzheimer's disease. Biomedical Signal Processing and Control, 88, 105669.

[13] Pena, D., Suescun, J., Schiess, M., Ellmore, T. M., Giancardo, L., & Alzheimer's Disease Neuroimaging Initiative. (2022). Toward a Multimodal Computer-Aided Diagnostic Tool for Alzheimer's Disease Conversion. Frontiers in neuroscience, 15, 744190.

[14] Rajalingam, B., Al-Turjman, F., Santhoshkumar, R., & Rajesh, M. (2022). Intelligent multimodal medical image fusion with deep guided filtering. Multimedia Systems, 28(4), 1449-1463.

[15] Rani, V. A., & Lalithakumari, S. (2019). Recent medical image fusion techniques: a review. Indian Journal of Public Health Research & Development, 10(7), 1399-1403.

[16] Singh, S., Singh, H., Bueno, G., Deniz, O., Singh, S., Monga, H., ... & Pedraza, A. (2023). A review of image fusion: Methods, applications and performance metrics. Digital Signal Processing, 137, 104020.

[17] Tawfik, N., Elnemr, H. A., Fakhr, M., Dessouky, M. I., & Abd El-Samie, F. E. (2021). Survey study of multimodality medical image fusion methods. Multimedia Tools and Applications, 80, 6369-6396.

[18] Tufail, A. B., Anwar, N., Othman, M. T. B., Ullah, I., Khan, R. A., Ma, Y. K., ... & Hamam, H. (2022). Early-stage Alzheimer's disease categorization using PET neuroimaging modality and convolutional neural networks in the 2D and 3D domains. Sensors, 22(12), 4609.

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[20] Venugopalan, J., Tong, L., Hassanzadeh, H. R., & Wang, M. D. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. Scientific reports, 11(1), 3254.

[21] Xia, K. J., Yin, H. S., & Wang, J. Q. (2019). A novel improved deep convolutional neural network model for medical image fusion. Cluster Computing, 22, 1515-1527.

[22] Yadav, S. P., & Yadav, S. (2020). Image fusion using hybrid methods in multimodality medical images. Medical & Biological Engineering & Computing, 58(4), 669-687.

[23] Zhang, H., Xu, H., Tian, X., Jiang, J., & Ma, J. (2021). Image fusion meets deep learning: A survey and perspective. Information Fusion, 76, 323-336.

[24] Zhou, T., Liu, M., Thung, K. H., & Shen, D. (2019). Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. IEEE transactions on medical imaging, 38(10), 2411-2422.